Template from www.lineupx.com/job-descriptions/ai-evaluation-analyst. Replace every highlighted [field] and keep the version you need: the full-time job for freshers or the internship.

AI evaluation analyst (fresher)

Company: [Company name]
Location: [Location] | Work mode: [Work mode]
Employment type: Full-time | CTC: [CTC for the full-time role]
Joining: [Joining date] | Apply by: [Apply by]

About [Company name]

[About your company]

About the role

[Company name] is hiring a fresher AI evaluation analyst for the [Team] team. You will rate AI answers against written guidelines, check claims against sources and record recurring errors. You will help maintain test questions so the team can compare results after changes. A senior reviews your ratings. Content reviewed: [ordinary product text only / describe potentially distressing content]. Support arrangements: [state available breaks, the escalation contact and how to hand back a batch, where relevant].

What you will do

- Read the written guidelines and their worked examples before each batch, and ask the guideline owner when a case is not covered
- Rate [chatbot answers / search results and summaries / translations (add the languages to the requirements) / extracted fields] on the scale in the guidelines, and write a one- or two-line reason for each rating
- Compare each factual claim with the source supplied in the task, or with the source the answer cites, and mark it supported, unsupported or contradicted
- Write test prompts for a feature: plain requests, vague ones, ones outside its remit and ones built to trip it up, then run them only on the test version named in the task
- Label and check text data against the labelling rules, and re-check a sample of your own labels the next day to see whether you still agree with them
- Log each error in the shared sheet with its type, such as an invented fact, a wrong figure, an ignored instruction or a mixed-up source, and keep one example per type for the engineers
- Help maintain the team's fixed set of questions with known-good answers: add new cases, retire outdated ones and rerun the set after each change to the model or its instructions, recording the result in the tracker
- Work through a shared sample with a second reviewer, list the cases where your ratings differ and bring them to the weekly calibration discussion
- Flag guidelines that are unclear, clash with one another or miss a case, and quote the example that showed the gap
- Use only the fictional or approved data in the task, and tell your senior at once if a batch holds real customer details or upsetting material you were not told about

What we are looking for

- Strong reading and writing in English: can read a long answer closely and explain a rating in one or two clear lines
- Careful judgement: applies a written guideline to each case in the same way, and says when a case does not fit it
- Spreadsheet skill in Excel or Google Sheets: filters, sorting, simple formulas and counts, and a tidy results sheet

- Honesty about uncertainty: writes ‘not sure’ with a reason and what would settle it, in place of guessing to finish the sheet
- The habit of checking a claim against its source, such as opening the page an answer cites and finding the sentence
- Patience for detailed, repeated checking, and the discipline to keep the same standard from the first item to the last
- An example of careful checking that you can describe or show, such as proofreading, fact-checking, marking a test or labelling data, from study, a project or voluntary work

Good to have

- Basic Python or SQL for opening data files and counting results
- Reading and writing Hindi or another Indian language, for checking answers given in it [move to requirements if the post rates answers in it]
- Knowledge of the subject the product covers, such as health, finance, education or software
- Time spent with a labelling or review tool, for example Label Studio, or in proofreading and fact-checking work
- A plain-language idea of what a large language model is and why it can state a wrong answer in a confident tone

Who can apply

Final-year students and graduates from any stream can apply, including BA, B.Com, B.Sc, BBA, BCA and B.Tech. Careful reading, clear writing and sound judgement matter more than the degree, and knowledge of the subject our product covers is a bonus.

How to apply

Apply through this job post by [Apply by]. Shortlisted candidates will hear from us about the next steps.

AI evaluation analyst intern

Company: [Company name]
Location: [Location] | Work mode: [Work mode]
Employment type: Internship | Duration: [Internship duration] | Stipend: [Stipend for the internship]
Joining: [Joining date] | Apply by: [Apply by]

About [Company name]

[About your company]

About the role

[Company name] is looking for an AI evaluation intern to rate and check model answers with a mentor, using fictional or approved material only. You will learn the guidelines, rate practice batches, record what goes wrong and help extend a set of test questions. If you wish to publish your work or show it in a portfolio, [Company name] has to approve it first.

What you will do

- Rate a practice batch of fictional answers, then compare your ratings with your mentor's and write down each difference and why it happened
- Check ten answers against a one-page source and mark each claim supported, unsupported or contradicted, with your mentor reading the result
- Write a set of test questions on one topic, each with the answer it should get and the page that proves it
- Group errors from a practice batch by type, count the items in each group and show your mentor one example of the most frequent error.
- Rewrite vague test questions so that a reviewer could mark the reply right or wrong, and show the before and after
- Note every place where the guidelines were unclear to you and suggest clearer wording with an example
- Present a short write-up of what you found to the team at the end

What we are looking for

- A final-year student or recent graduate, in any stream
- Reads English closely and can explain a judgement in two or three clear sentences
- Can sort, filter and count in a spreadsheet
- Willing to follow written guidelines and to write 'not sure' when a case is unclear
- Can describe one time you checked work or facts carefully, such as proofreading, marking a test or comparing a claim with its source

Who can apply

Final-year students and graduates from any stream can apply, including BA, B.Com, B.Sc, BBA, BCA and B.Tech. Careful reading, clear writing and sound judgement matter more than the degree, and knowledge of the subject our product covers is a bonus.

How to apply

Apply through this job post by [Apply by]. Shortlisted candidates will hear from us about the next steps.

Screening questions to add to your job post

1. **An AI answer sounds confident, but the source it cites does not say that. What do you do?** Look for opening the source, finding the sentence or its absence, marking the claim unsupported and noting the type of error. Accept a written note or a flag in a tool. The aim is that nothing passes because it sounds right.
2. **Two reviewers rate the same answer differently. How do you settle it?** Look for going back to the guideline wording first, each person giving a reason, and a question to the guideline owner when the text does not decide the case. Accept a short call or a written exchange.
3. **What makes an evaluation guideline unclear? Give one example from a marking scheme, rubric or checklist you have used.** Look for words open to different readings, such as ‘good’ or ‘helpful’, a case the guideline misses, or two rules that clash. Accept an example from coursework, a quiz key or any other checklist.
4. **A fictional policy says, ‘Employees need their manager’s written approval before booking leave.’ Rewrite ‘Is the leave policy good?’ as one question whose answer can be checked against that sentence.** Look for a question such as ‘Whose written approval is needed before booking leave?’ Accept any clear question answerable from the supplied sentence, without invented policy details.
5. **You cannot tell whether an answer is right. What do you put in the sheet?** Look for ‘not sure’ with the reason and what would settle it, such as a source or a subject expert. Accept flagging the item for a senior’s review.
6. **An answer is polite and well written but gets a date wrong. How do you rate it, and what reason do you write?** Look for correctness judged apart from style, and a reason that names the error. Accept any scale, as long as the reason says which part earned which score.
7. **In a sheet of rated answers, how would you count the errors by type?** Look for an error-type column and a filter, pivot table or count formula. Accept any equivalent method.